

IN THE CLAIMS

This listing of claims will replace all prior versions and listings of claims in the application.

1. (Currently Amended) A method for automatically evaluating an essay to detect at least one writing style error, excluding misspellings, comprising:
electronically receiving an essay on a computer system;
assigning a feature value for each of one or more features for one or more text segments in the essay, wherein the feature values are automatically calculated by the computer system and wherein function words of the essay are excluded when calculating the feature values;
storing the feature values for the one or more text segments on a data storage device accessible by the computer system;
comparing the feature values for each text segment with a model configured to identify at least one writing style error, wherein the model includes at least one decision tree to determine a probability associated with a likelihood of the at least one writing style error, and wherein the at least one decision tree is generated based on at least one human evaluated essay; and
displaying an indication of an identified writing style error.
- 2-3. (Canceled)
4. (Previously Presented) The method of claim 1 wherein the comparison step comprises extracting patterns from the feature values, wherein the patterns are based on the presence or absence of features associated with each word in the essay.
- 5-6. (Canceled)
7. (Original) The method of claim 1 wherein the feature values comprise the ratio of the evaluated text segment occurrences in the essay to the total number of text segments in the essay.

8. (Previously Presented) The method of claim 1 wherein the feature values comprise the average, over all paragraphs of the essay, of the ratio of the number of times the evaluated text segment occurs in a paragraph of the essay, over the total number of text segments in the paragraph.

9. (Previously Presented) The method of claim 1 wherein the feature values comprise the largest value of the ratio of the number times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph, wherein the ratio is calculated for each paragraph in the essay.

10. (Original) The method of claim 1 wherein the feature values comprise the length, measured in characters, of the text segment.

11. (Original) The method of claim 1 wherein the feature values comprise a value indicating whether the text segment includes a pronoun.

12. (Currently Amended) The method of claim 1 wherein the feature values comprise a value representing the interval distance between consecutive text segment occurrences determined by calculating the number of intervening characters.

13-15. (Canceled)

16. (Currently Amended) A system for automatically evaluating an essay to detect at least one writing style error, excluding misspellings, comprising:

a computer system configured to electronically receive an essay;

a feature extractor configured to assign a feature value for each of one or more features for one or more text segments in the essay wherein function words of the essay are excluded when calculating the feature values;

a data storage device, connected to the computer system, configured to store the feature values for the one or more text segments;

a feature analyzer configured to evaluate the essay for at least one writing style error by comparing the feature values for each of one or more text segments with a model generated using at least one human evaluated essay, wherein the model includes at least one decision tree to determine a probability associated with a likelihood of the at least one writing style error, and wherein the at least one decision tree is generated based on at least one human evaluated essay; and

a display for presenting the evaluated essay, wherein the evaluated essay includes an indication of at least one identified writing style error.

17-20. (Canceled)

21. (Original) The system of claim 16 wherein the feature extractor comprises an essay ratio calculator configured to generate a value representing the ratio of the number of times the evaluated text segment occurs in the essay to the total number of text segments in the essay.

22. (Original) The system of claim 16 wherein the feature extractor comprises an average paragraph ratio calculator configured to generate a value representing the average over all paragraphs in the essay of the ratio of the number of times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

23. (Original) The system of claim 16 wherein the feature extractor comprises a highest paragraph ratio calculator configured to generate a value representing the largest ratio of the number of times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

24. (Original) The system of claim 16 wherein the feature extractor comprises a length calculator configured to generate a value representing the length, measured in characters, of the text segment.

25. (Original) The system of claim 16 wherein the feature extractor comprises an identifier to determine whether the text segment includes a pronoun.

26. (Currently Amended) The system of claim 16 wherein the feature extractor comprises a distance calculator configured to generate a value representing the distance measured in characters between consecutive text segment occurrences.

27-30. (Canceled)

31. (Withdrawn) A method for generating a model for determining overly repetitive text segment use, comprising:

electronically receiving training data on a computer system wherein the training data comprises an essay annotated to identify one or more text segments used in an overly repetitive manner;

assigning a feature value for each of one or more features for each text segment in the essay, wherein the feature values are automatically calculated by the computer system;

assigning an indicator value for each text segment in the essay, wherein the indicator value is set at a first value and if the text segment has been used in an overly repetitive manner;

storing the feature values and the indicator value for each text segment in the essay in a data storage device accessible by the computer system; and

creating a model for overly repetitive use of the one or more text segments in the essay by identifying patterns in the feature values wherein the patterns are identified by a machine learning tool.

32. (Withdrawn) The method of claim 31 wherein the text segment comprises a word.

33. (Withdrawn) The method of claim 31 wherein the annotations are manual markings.

34. (Withdrawn) The method of claim 31, wherein the function words of the essay are not considered by the computer system in calculating the feature values.

35. (Withdrawn) The method of claim 31 wherein the feature values comprise the total number of times the evaluated text segment occurs in the essay.

36. (Withdrawn) The method of claim 31 wherein the feature values comprise the ratio of the evaluated text segment occurrences in the essay to the total number of text segments in the essay.

37. (Withdrawn) The method of claim 31 wherein the feature values comprise the average over all paragraphs of the essay of the ratio of the number times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

38. (Withdrawn) The method of claim 31 wherein the feature values comprise the largest value of the ratio of the number times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph, wherein the ratio is calculated for each paragraph in the essay.

39. (Withdrawn) The method of claim 31 wherein the feature values comprise the length, measured in characters, of the text segment.

40. (Withdrawn) The method of claim 31 wherein the feature values comprise a value indicating whether the text segment includes a pronoun.

41. (Withdrawn) The method of claim 31 wherein the feature values comprise a value representing the interval distance between consecutive text segment occurrences.

42. (Withdrawn) The method of claim 41 wherein the distance is determined by calculating the number of intervening words.

43. (Withdrawn) The method of claim 41 wherein the distance is determined by calculating the number of intervening characters.

44. (Withdrawn) A system for generating a model useful in determining overly repetitive text segment use, comprising:

a computer system configured to receive training data, wherein the training data comprises an essay annotated to identify one or more text segments used in an overly repetitive manner;

a feature extractor configured to calculate a feature value for each of one or more features for each text segment in the essay and to assign an indicator value for each text segment in the annotated essay, wherein the indicator value indicates whether the text segment has been used in an overly repetitive manner;

a data storage device configured to store the feature values and the indicator value for each text segment in the essay;

a machine learning tool configured to analyze the features to identify patterns; and

a model builder to create a model for overly repetitive use of the text segments, wherein the model is constructed from the identified patterns.

45. (Withdrawn) The system of claim 44 wherein the annotated essays are manually marked.

46. (Withdrawn) The system of claim 44 wherein the feature extractor comprises an occurrences calculator configured to generate a value representing the total number of times the text segment occurs in the essay.

47. (Withdrawn) The system of claim 44 wherein the feature extractor comprises an essay ratio calculator configured to generate a value representing the ratio of the number of times the evaluated text segment occurs in the essay to the total number of text segments in the essay.

48. (Withdrawn) The system of claim 44 wherein the feature extractor comprises an average paragraph ratio calculator configured to generate a value representing the average over all paragraphs in the essay of the ratio of the number of times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

49. (Withdrawn) The system of claim 44 wherein the feature extractor comprises a highest paragraph ratio calculator configured to generate a value representing the largest ratio of the number of times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

50. (Withdrawn) The system of claim 44 wherein the feature extractor comprises a length calculator configured to generate a value representing the length, measured in characters, of the text segment.

51. (Withdrawn) The system of claim 44 wherein the feature extractor comprises an identifier to determine whether the text segment includes a pronoun.

52. (Withdrawn) The system of claim 44 wherein the feature extractor comprises a distance calculator configured to generate a value representing the distance between consecutive text segment occurrences.

53. (Withdrawn) The system of claim 52 wherein the distance between consecutive text segment occurrences is measured in words.

54. (Withdrawn) The system of claim 52 wherein the distance between consecutive text segment occurrences is measured in characters.